

SZINUSZHULLÁMÚ ÉS AMPLITÚDÓMODULÁLT BESZÉD ÉSZLELÉSÉNEK VIZSGÁLATA MAGYAR MONDATOK SEGÍTSÉGÉVEL

JAKAB ZOLTÁN¹ – NAGYNÉ RINGER GABRIELLA² – VÍG JULIANNA¹ –
SZABÓ PÁL TAMÁS³

¹Eötvös Loránd Tudományegyetem, Bárczi Gusztáv Gyógypedagógiai Kar,

²Pedagógiai Szakszolgálat, Budapest, XVII. kerület,

³Szent János Kórház és Észak-Budai Egyesített Kórházak, Neurológiai Osztály

E-mail: jakab.zoltan@barczy.elte.hu

Beérkezett: 2017. október 18. – Elfogadva: 2018. január 20.

Online megjelenés: 2018. május 17.

Háttér és célok: Kutatásunkban elszegényített beszédíngerek két típusát használtuk gyerekek és felnőttek beszédészlelési képességének vizsgálatára.

Módszer: Vizsgálatunkat 5-6 éves gyermek és felnőtt mintán végeztük, szinuszhullámú és amplitúdómodulált beszédíngerek – magyar nyelvű mondatok – segítségével. Beszédíngereinket természetes kiejtésű változatokból szintetizáltuk.

Eredmények: Korábbi angol nyelvű vizsgálatok eredményeihez hasonlóan azt tapasztaltuk, hogy mindkét korcsoport igen jól érti ezeket a szintetizált beszédíngereket, bár a felnőttek jobban teljesítenek a gyerekeknél. Ugyanakkor adataink eltérő képet mutatnak a kétféle szintetizált beszédíngertípus viszonyáról az angol nyelvű vizsgálatokhoz képest. Míg angol anyanyelvű mintáknál a gyerekek a szinuszhullámú beszédet relatíve jobban értették, mint az amplitúdómoduláltat, felnőtteknél pedig ez a viszony megfordult, a magyar nyelvű vizsgálatok esetén gyerekeknél egyforma teljesítményt találtunk a kétféle íngertípusra, míg felnőtteknél az amplitúdómodulált íngerek előnyét találtuk (lehetséges ugyanakkor, hogy ez csak az íngerehészési paraméterek eltérő beállításának következménye). A lexikai restaurációs hatásokat jelezte, hogy a mondatok első főnévi csoportjának felismerése igen erősen korrelált a mondatok további részének felismerésével.

Konklúzió: Adataink részben reprodukálják az angol nyelven végzett vizsgálatok eredményeit. Új eredményként kétféle perceptuális tanulási hatást sikerült elkülönítenünk: az azonnali perceptuális tanulás gyerekeknél és felnőtteknél is megfigyelhető, míg a próbák során jelentkező gyakorlási, tehát az egyes mondatok közötti transzferhatás csak felnőtteknél.

Kulcsszavak: szinuszhullámú beszéd, amplitúdómodulált beszéd, beszédészlelés, perceptuális tanulás

BEVEZETÉS

A beszédészlelés vizsgálatában számos adat, megfigyelés utal arra, hogy az emberi beszédet nehezített körülmények között is igen jól meg tudjuk érteni. Ennek egyik alapvető példája a zajos környezetben hallható beszéd (Nilsson, Soli és Sullivan, 1994; Nilsson, Soli és Gelnett 1996; Gósy, 1995). Elméleti szempontból talán még érdekesebb eset az, amikor több személy egyidejű beszédéből szűrjük ki és követjük az egyikük hangját (ez az ún. koktélparti helyzet, lásd Cherry, 1953). A koktélparti jelenség azt is jelzi, hogy képesek vagyunk dinamikusan változtatni a jel és a zaj kijelölését és az ennek alapján való szűrést a beszédészlelés során. Ez pedig felveti azt a kérdést, hogy melyek azok az ingerjellemzők, melyek alapján a beszédet beszédként ismerjük föl, és a benne lévő különböző szegmenseket kiemeljük, azonosítjuk. Utóbbi kérdés az elszegényített szintetikus beszédű vizsgálatok útján közelíthető meg; ezeknek két közismert formája a szinusz hullámú beszéd (*sine-wave speech*, SWS) és az amplitúdómodulált beszéd (*vocoded speech*, VOC, illetve más helyeken *amplitude envelopes*, AE; Remez, Rubin, Pisoni és Carrell, 1981; Shannon, Zeng, Kamath, Wygonski és Ekelid, 1995). Mára mindkettőnek számos variánsa létezik, de eredeti formában a szinusz hullámú beszéd elsősorban a természetes beszéd formánsainak időben lassabb frekvenciaváltozásait (az ún. *dinamikus spektrális struktúrát*; Nittrouer és Lowenstein, 2010) őrzi meg, míg az amplitúdómodulált beszéd épp ezeket mossa el kisebb-nagyobb mértékben, és a természetes beszéd amplitúdóváltozásait őrzi meg, néhány előre kijelölt frekvenciasávban.¹ Mindkét típusú beszéd szintézist használják a beszédészlelés vizsgálatában; mára angol nyelvű anyagokon angol anyanyelvű kísérleti személyekkel számos vizsgálat született (Remez és mtsai, 1981; Remez, Rubin, Nygaard és Howell, 1987; Remez, Rubin, Berns, Pardo és Lang, 1994; Remez és Rubin, 1990; Shannon és mtsai, 1995; Rosner és mtsai, 2003; Nittrouer, Lowenstein és Packer, 2009; Nittrouer és Lowenstein, 2009, 2010; Hervais-Adelman, Davis, Johnsrude, Taylor és Carlyon, 2011; Hillenbrand, Clark és Baer, 2011; Lowenstein, Nittrouer és Tarr, 2012; Newman, Chatterjee, Morini és Nasuta, 2013). Létezik ugyanakkor néhány nyelvközi összehasonlítás is, ezek az angol mellett német, illetve mandarin kínai nyelven gyűjtöttek adatokat (Bent, Loebach, Phillips és Pisoni, 2011; Rosen és Hui, 2015).

Klasszikus beszédészlelési vizsgálatok szinusz hullámú beszéddel

A beszédészlelés az inger összetett természetéből fakadóan hierarchikus folyamat, melyben az akusztikai jellemzők elemzését követi a fonetikai jegyek azonosítása, majd a nyelvspecifikus szabályokon alapuló fonémadöntés (Gósy, 2005; Johnson, 2010; Remez, 2005). Már az 1980-as években is ismert volt, hogy míg a fonetikai kategorizáció

¹ Kissé árnyalja ezt a képet, hogy a szinusz hullámú beszéd előállítására szolgáló algoritmusok általában meghagyják a formánsok amplitúdóváltozásainak egy részét, tehát a szinusz hullámmal közelített formánsok amplitúdója egy-egy ilyen ingeren belül nem konstans. Legalábbis alapesetben nem az, de a szintézis algoritmusának megfelelő beállításával az amplitúdók konstanssá tehetőek (Remez és Rubin, 1990).

számos egyszerűbb jelensége kiváltható nem beszédjellegű ingerekkel, a fonetikai feldolgozás összetettebb jelenségei már csak beszédingerek esetében figyelhetőek meg. Repp (1982) például amellest érvel, hogy a beszédjellegű és nem beszédjellegű ingerek feldolgozására két elkülönült feldolgozási módot használunk, és léteznek olyan ingerek, melyek megfelelő kontextusban bemutatva e kettő közül bármelyiket, illetve akár egyszerre mindkettőt is aktiválják. A szinuszhullámú beszéd különösen alkalmas ennek a jelenségnek a vizsgálatára; ugyanis naiv hallgatók ezt nem automatikusan kezelik beszédingerként, csak akkor, ha erre előzetes instrukcióval felhívják a figyelmüket. Sőt, az ún. *fonetikai kölcsönhatások*² egyazon inger esetében is megfigyelhetőek akkor, ha azt beszédként észleljük, és nem figyelhetőek meg, ha a hallgató nem ismeri fel, hogy beszédről van szó. Ezekhez szorosan kapcsolódó, és ugyancsak a kétféle percepció mechanizmus meglétét támogató evidencia az ún. „duplex percepció”, mely a vizuális modalitásban megjelenő multistabil percepció auditoros megfelelője (Repp, 1982; Remez és mtsai, 1994; Schwartz, Grimault, Hupe, Moore és Pressnitzer, 2012). A duplex percepció jelensége azt jelenti, hogy egy ingert megjelenésekor egyszerre hallunk beszédnek és nem beszédjellegű ingernek. Ilyen észlelés figyelhető meg, ha például egy szintetizált szótag egyik akusztikus összetevőjét (például a formánsátmenetet) az egyik fülbe, míg a többi összetevőt a másik fülbe játsszuk le. Ilyenkor a személy két dolgot hall egyszerre: a szótag egészét, és ezzel egy időben a formánsátmenet ciripelésre emlékeztető hangját.

Felvetődhet a kérdés, hogy ezek a feldolgozási szintek a hallórendszeren (illetve nyelvi rendszeren) belül mennyiben felelnek meg a mentális modularitás kritériumainak, tehát mennyiben tekinthetőek elkülönült moduloknak. Remez és munkatársai (1994) szerint a két feldolgozási szint egyaránt gyors és területspecifikus annyiban, hogy eltérő perceptuális integrációs elvek szerint működik. Remez és munkatársai ugyanis amellest érvelnek, hogy a hallásijelenet-elemzés (Bregman, 1990; magyarul lásd Honbolygó, 2007a, 330) elvei nem magyarázzák meg a szinuszhullámú beszéd észlelésének jelenségét, azt tehát, hogy három, időben egymástól függetlenül változó frekvenciájú szinuszhullámot koherens beszédként vagyunk képesek észlelni. A beszédészlelés során fellépő perceptuális integráció más elvek szerint működik. Másfelől viszont a szinuszhullámú beszéd észlelésének instrukcióérzékenysége azt mutatja, hogy a beszédészlelésért felelős mechanizmusok nem automatikusak, és nem is magukba zártak. Tehát feldolgozási szintekről beszélhetünk, ám fodori értelemben vett modulokról kevésbé. Kifinomult kísérleti technikákkal e feldolgozási szintek tovább is tagolhatóak. Így például Samuel és Kat (1996) a szintetikus /ba-/da/ kontinuumban megfigyelhető adaptációs hatásokat vizsgálva összetett kritériumot használt: (i) az adaptáció csak monaurálisan avagy interaurálisan is megfigyelhető-e, és (ii) hogyan jelenik meg az adaptáció a kategóriahatár-változásban, illetve a reakcióidőben. Kísérlet-sorozatuk alapján a szerzők három feldolgozási szintet javasoltak: (1) egy monaurális

² A 'phonetic trading relations' terminus fordítása (Repp, 1982). Szinte mindegyik fonetikai kontraszt (pl. szár-sztár) esetén több akusztikai jegy is segíti a megkülönböztetést. Az egyes akusztikus jegyek egymás hatását erősíteni, illetve ellensúlyozni is képesek. Például a szár-sztár pár esetében az /sz/ hang utáni megnövekedett szünet valószínűbbé teszi, hogy 'sztár'-t hallunk, ám azt bizonyos határon belül ellensúlyozni lehet az /á/ hang első formánsa (F1) frekvenciájának megváltoztatásával (például szintetizált beszédben; Morrongiello, Robson és Best, 1984).

akusztikus elemző szintet (ezt nemfonetikus adaptorok hajtották meg, és csak reakcióidő-változást eredményezett a /ba/-/da/ döntésben); (2) egy binaurális, integratív-akusztikus szintet: ezt is nemfonetikus adaptorok befolyásolták, de itt az adaptáció kontralaterálisan is megfigyelhető volt, és nemcsak a reakcióidőt lassította, hanem a kategóriahatárt is eltolta; végül (3) egy binaurális fonetikus szintet, melyre csak fonetikus adaptorok hatottak, és ahol az adaptáció csak a kategóriahatár változásában mutatkozott meg.

Ebbe a tágabb elméleti keretbe illeszkedett tehát Remez és munkatársai számos vizsgálata az 1980-as, 1990-es években. A beszédészlelés elkülönült feldolgozási szintjét véleményük szerint az is alátámasztotta, hogy a szinuszbeszédet annak ellenére értjük – s ebben az értelemben koherensen észleljük –, hogy az inger akusztikus struktúráját a hallásijelenet-elemzés elvei nem képesek egységes mintázattá integrálni (Remez és mtsai, 1994; magyarul lásd Honbolygó, 2007b, 369). A beszédészlelés során fellépő perceptuális integrációnak tehát más elvek szerint kell haladnia. Ilyen elv lehet például, hogy a három szinuszformáns szinte a valóságosnál is kontrasztosabban, egyértelműbben jeleníti meg az eredeti mondat globális formánsstruktúráját; ez fontos támpont lehet a megértésben. Egészében tehát a szinuszhullámú beszéd „beszédszerű” idői mintázatot mutat, annak ellenére, hogy a formánsok felharmonikusai és a mássalhangzók zörejkomponenseinek túlnyomó többsége hiányzik belőle. A három szinuszformáns mellett ugyanakkor fontos szerepe van a felismerhetőségben a zárhangoknak megfelelő szüneteknek, amelyek alatt a szinuszformánsok amplitúdója is minimálisra csökken (Remez és Rubin, 1990). Szintén a beszédfeldolgozás sajátosságait mutatja szinuszbeszéd esetében a forrásnormalizáció (speaker normalization) megléte (Remez és mtsai, 1987); a klasszikus, Ladefoged és Broadbent (1957) által végzett „/bit/-/bet/-/bat/-/but/” kísérlet ugyanis szinuszhullámú ingerekkel részben megismételhetőnek bizonyult, és az eredmények alapján levonható a következtetés, hogy normalizációs hatások szinuszhullámú beszéd esetén is működnek.

Szintetizált beszéd: fejlődési jelenségek

A beszédészlelés fejlődésével kapcsolatos vizsgálatok egyik felvetése az, hogy a különböző akusztikus ingerjegyek súlyozása a fonetikai döntésekben jellegzetes életkori változásokat mutat, és ez összefügghet a befogadó anyanyelvével, az abban való korai tapasztalatszerzéssel (Nittrouer és Lowenstein, 2009). Két ilyen jellegzetes ingervonás a formánsátmenetek, illetve a mássalhangzók zörejkomponensei. Ezek súlyozása a beszédészlelési feldolgozásban különbözik kisgyerekek és felnőttek között. Nittrouerék adatai szerint három- és ötévesek nagyobb mértékben hagyatkoztak a formánsátmenetekre, és kisebb mértékben a zörejkomponensekre, mint hétévesek és felnőttek, ha CV szótagok azonosítása a feladat. E jelenségek vizsgálata a természetes beszéd észlelése mellett szinuszhullámú, amplitúdómodulált, illetve egyéb módon nehezített (pl. suttogva kiejtett) beszédingerekkel történhet. Nittrouerék e vizsgálatában a természetes beszéd mellett szinuszhullám-szintézist és természetes suttogó beszédet használtak az ingerként használt szótagok előállítására. Eredményeik szerint öt- és hétévesek akkor is kitarítottak a formánsátmenetekre alapozott fonémadöntés mellett, ha

a formánsátmeneteket szinuszhullámú szintetizált formában hallották, míg felnőttek ebben az esetben kevésbé vették figyelembe a formánsátmeneteket, mint a természetes kiejtésben. A természetes kiejtés, illetve suttogó beszéd összevetése ezzel ellentétes eredményt adott: itt a felnőttek mindkét esetben egyenlő mértékben hagyatkoztak a formánsátmenetre, míg a gyerekek (5 és 7 évesek) sokkal kevésbé használták a formánsátmeneteket a fonetikai kategorizációban, ha zöngés természetes kiejtésről suttogó beszédre váltottak az ingerek. Ez az eredmény számos fontos megfigyelést foglal magában. Egyrészt a szinuszhullámú formánsok egyáltalán nem hangzanak természetesen a gyerekek számára, mégis erősen használják őket a fonémadöntésben, míg a suttogó beszédben megjelenő elmosódottabb, zaj jellegű formánsokat annak ellenére sem használják, hogy a suttogó beszéd hangzása ismerős lehet számukra. Mind a szinuszhullám-szintézis, mind a suttogás eliminálja a természetes, zöngés formánsok egyik alapvető akusztikus tulajdonságát, nevezetesen a harmonikus szerkezetet (a felharmonikusok sorát, mely a beszélő hangszínének meghatározásában is fontos szerepet játszik). Bár a harmonicitást mindkét transzformáció eltünteteti, mégis az egyik esetben a gyerekek kitaranak a formánsátmenetekre való fókusz mellett, a másikban nem. A különbség oka az lehet Nittrouerek szerint, hogy míg a szinuszbeszédben a formánsok élesen elkülönülnek egymástól, így frekvenciáik időbeli változása jól követhető (tulajdonképpen még a természetes beszédnél is jobban), addig a suttogásban a formánsok csak elmosódott, zajszerű formában jelennek meg. A gyerekek pedig, értelmezésük szerint, a jól követhető, időben lassabb, dinamikus frekvenciamodulációra érzékenyebbek, míg a zajszerű elemekre kevésbé. Ezzel szemben a felnőttek jobban fókuszálnak a zajszerű ingerjegyekre, és valamivel kevésbé a frekvenciaváltozásokra. Ezzel a következtetéssel Nittrouerek lényegében csatlakoznak Remez és munkatársai kritikájához a hallásijelenet-elemzéssel kapcsolatban. Ugyanis a bregmani felfogás szerint a harmonikus struktúra a hallási perceptuális integráció egyik alapvető vezérfonala, míg Nittrouerek adatai szerint a szótagok észlelésében egyáltalán nem játszik meghatározó szerepet. Későbbi vizsgálataikban a szerzők (Nittrouer és mtsai, 2009; 2010; Lowenstein és mtsai, 2012) ezt az alapkövetkeztetést értelmes és értelmetlen mondatok vizsgálatával, illetve a szintézis módjának változtatásával is megerősítették. Létezik ennek a tendenciának ellentmondó adat is, 27 hónapos gyerekek vizsgálatából, akik még nem mutatták a szinuszhullámú beszéd felismerési előnyét az amplitúdómodulálttal szemben (Newman és mtsai, 2013). Ezt az eltérést az utóbbi vizsgálat szerzői módszertani különbségekkel magyarázták (pl. az ő feladatukban egy célszó hatására kellett a tipegő korú személyeknek a két kép közül a megfelelőre nézni). Ez azonban véleményünk szerint nem tűnik kielégítő magyarázatnak,³ így felvetődhet, hogy két- és hároméves kor között is vannak életkori különbségek a beszédészlelési stratégiákban.

E fejlődési megfontolások felvetik azt a kérdést is, hogy az egyes nyelvek nem különböznek-e a súlyozási stratégiák tekintetében. Nittrouerek szerint igen: a gyerekeknek a nyelvvelajátítás korai időszakában, lényegében már csecsemőkortól kezdve rá kell

³ A Nittrouer-féle szótagos vizsgálatban (Nittrouer és Lowenstein, 2009) zárt válaszkészletet használtak (mindössze négy szótag közül lehetett választani); ugyanebben a vizsgálatban nem volt kontextushatás sem. Ennek ellenére ellentétes eredményt kaptak, mint Newmanék (2013). Más magyarázatra van tehát szükség; lehet, hogy 2 és 3 évesek között van valamilyen életkori különbség.

találniuk azokra az észleléses jegyekre, amelyek az adott anyanyelvi környezetben a legrelevánsabbak a jelentés és nyelvtani szerkezet rekonstrukciója szempontjából, és ezeket a jegyeket fogják a legalaposabban feldolgozni, vagyis nagyobb súllyal fogják figyelembe venni, mint a kevésbé releváns vonásokat (Nittrouer és Lowenstein, 2009). Ugyanakkor a súlyozási stratégiák nyelvek közti különbségeiről minimális adat áll csak rendelkezésre, magyar nyelven pedig egyáltalán nincs még ezzel kapcsolatos vizsgálat. Ez motivációt jelenthet arra, hogy a későbbiekben e fenti vizsgálatokhoz hasonlókat magyar nyelven is elvégezzünk.

Egy ehhez szorosan kapcsolódó kérdés az angol és a magyar nyelv közti hangtani különbségeknek a megértésre gyakorolt hatása. A magyar és az angol nyelv kiejtése, illetve hangrendszere között számos olyan különbség van, melyek a beszédértés, és különösen a szinusz hullámú vagy amplitúdómodulált beszéd megértése szempontjából kritikusak. Például az angolban bizonyos magánhangzók, illetve kettőshangzók kiejtésének hosszát befolyásolja hangsúlyos szótagban a közvetlenül utánuk következő mássalhangzó zöngéssége: ugyanazt a magán- vagy kettőshangzót hosszabban ejtik zöngés, mint zöngétlen mássalhangzó előtt. Például rövid magánhangzók megnyúlnak szövégi zöngés mássalhangzó előtt: 'sad', 'lab', zöngétlen előtt azonban nem változnak: 'sat', 'lap'. Magyarban nincs ilyen különbség: 'rak' vs. 'rag'; 'kéz' vs. 'kész', 'hat' vs. 'had'. Egy másik közismert példa a hangsúlyozás különbsége: az angolban, szemben a magyarral, nem mindig az első szótag hangsúlyos. Sőt, a hangsúlynak viszonylag szisztematikus jelentésmegkülönböztető szerepe is van, például ige-főnév párok esetén: / 'permit/ (főnév) vs. /per'mit/ (ige); / 'record/ (főnév) vs. /re'cord/; / 'present/ (főnév) vs. /pre'sent/ (ige). Egy harmadik, számunkra roppant fontos különbség, hogy az angol szavak kiejtés szempontjából általában „közelebb” vannak egymáshoz, mint a magyarok. Ez azt jelenti, hogy egyrészt az angolban több beszédhang van (főleg magánhangzóból), mint a magyarban. Ugyanakkor az angol szavak átlagosan rövidebbek, mint a magyarok, így angolban gyakrabban fordul elő, hogy ha egy szó valamely hangját kicsit másképp ejtjük, akkor akaratlanul is egy másik szóhoz jutunk.⁴ Későbbi kutatásokban az ilyen típusú különbségekre (azok megértéssel való kapcsolatára) vonatkozóan specifikus hipotéziseket is meg lehet fogalmazni. Jelen vizsgálatunk az első lépés a nyelvközi különbségek vizsgálata irányába, de hangsúlyoznunk kell, hogy itt most elsősorban a fejlődés egy másik aspektusára, nevezetesen a gyors perceptuális tanulásra voltunk kíváncsiak.

Azonnali perceptuális tanulás

A redukált beszédformák észlelésével, illetve ennek fejlődésével kapcsolatban felmerül a perceptuális tanulás kérdése is: vajon hogyan és milyen mértékben képesek különböző életkorú személyek teljesítményjavulásra. Ezzel kapcsolatban az egyik legérdekesebb jelenség a következő. Sok személy először hallván például egy szinusz hullámú mondatot, abból semmit nem ért; ha ezután meghallgatja ugyanazt a mondatot

4 Tömör magyar nyelvű forrás erről: http://idegennyelvor.blog.hu/2009/06/27/kulonbsegek_angol_es_magyar_kiejtese_kozott. Szakmailag alapos áttekintést ad a témáról Canepari, 2007.



1. ábra. Kéttónusú, ún. Mooney-kép és eredeti, tónusgazdag változata
(A szerzők saját felvétele)

természetes kiejtésben, majd ismét a szinuszhullámú változatot, akkor utóbbit már tökéletesen érti, és ezt egy „aha” élmény kíséri. Ennek a jelenségnek van egy vizuális párhuzama, és ennek kapcsán nevet is kapott: Craig M. Mooney kanadai pszichológus kéttónusú képeket mutatott kísérleti személyeinek, melyeket azok nem voltak képesek felismerni; viszont a képek tónusgazdag változatainak bemutatása után nyilvánvalóvá vált számukra, hogy mit ábrázolnak a kéttónusú képek (1. ábra; Mooney, 1957).

E jelenség fő vonásai a következők: a tónusgazdag kép egyszeri bemutatása teljesen megváltoztatta az eredeti, elszegényített inger észlelését, holott maga az eredeti inger egyáltalán nem változott. Ez a változás azonnali volt, erőfeszítést nem igényelt. A hatás a megfigyelések szerint tartós is volt, tehát akár évekkel a tónusgazdag kép első bemutatását követően is kimutatható volt. Mooney ezt a hatást azonnali perceptuális tanulásnak (instant learning; „one-shot” learning) nevezte el. Kovács Ilna és munkatársai egy 8 fős mintán alapuló vizsgálatukban úgy találták, hogy ez az azonnali tanulási hatás 4-5 éves gyerekeknél még nem figyelhető meg; ők akkor sem ismerik fel az elszegényített képet, ha egymás mellett mutatják be a tónusgazdag és az elszegényített képet (Kovács és Eisenberg, 2004). Kovács és munkatársai szerint e különbségnek két oka lehet: (i) az alacsony szintű perceptuális integráció csökkent mértéke, vagy (ii) a magas szintű integráció, a tárgyészlelési tapasztalatot igénylő perceptuális kategóriák (vagyis a top-down hatások) éretlensége. Erről a területről azonban hiányoznak a további kutatások. Annyit látunk tehát, hogy az azonnali perceptuális tanulás felnőtteknél mindkét modalitásban megjelenik, gyerekeknél pedig vizuális ingerek esetén valószínűleg nem. Mi tehát a helyzet a szinuszbeszéd észlelésével gyerekek esetében? Az előző részben idézett fejlődési adatokból úgy tűnik, hogy beszédingerek esetében gyerekeknél is többé-kevésbé megfigyelhető az azonnali tanulási hatás (Nittrouer és munkatársai 3 és 7 éves kor közötti életkori tartományban vizsgáloztak); a természetes kiejtés meghallgatása határozottan javította elsősorban a szinuszhullámú beszéd megértését gyerekeknél. A beszédészlelés fejlődésének jelen cikkben bemutatott vizsgálata pont ezt az aspektust vizsgálja, tehát hogy a redukált beszédformák észlelése során milyen különbségek figyelhetők meg a perceptuális tanulásban gyerekek és felnőttek között.

Összefoglalva tehát, vizsgálatunk célja az volt, hogy magyar ingeranyagon, kétféle redukált beszédinger segítségével hasonlítsuk össze gyerekek és felnőttek beszédészlelési teljesítményét. A magyar nyelven végzett vizsgálat értelme az lehet, hogy egy, az angoltól eltérő hangtani sajátosságokkal rendelkező nyelv esetén a kétféle szintézis eltérő hatással lesz a teljesítményre, és a különbségek értelmezhetőek is lesznek. A teljesítményen belül egyrészt az életkorral való általános javulásra voltunk kíváncsiak (ez az angol nyelven végzett vizsgálatokban is központi helyen állt), másrészt a perceptuális tanulás jelenségeire; ezt valamivel alaposabban elemeztük, mint a korábbi kutatások szerzői.

MÓDSZER

Személyek

Vizsgálatunkban 104 személy vett részt: 52 gyermek, akik életkora 60 és 72 hónap között volt (átlag: 67.2 hónap), illetve 52 felnőtt (életkori tartomány: 18–42 év; átlag: 23.92 év). A felnőtt- és gyermekcsoportot is két egyenlő részre osztottuk, így négy 26 fős almintát kaptunk. 26 gyermek és 26 felnőtt kapott szinuszhullámú beszédingereket (*gyerek szinusz* és *felnőtt szinusz* csoport), mindkét korcsoportból ugyanennyien pedig amplitúdómodulált ingereket (*gyerek VOC* és *felnőtt VOC* csoport). A két gyermekcsoport életkori adatai igen hasonlóak voltak (szinuszcsoport: 61–72 hónap, átlag 67.35 hónap; VOC csoport: 60–72 hónap, átlag: 67.08 hónap). Nemek szempontjából mind a négy almintá kiegészített volt: 13-13 lány/fiú, illetve nő/férfi vett részt a kísérletben. A felnőtt kísérleti személyek 18. életévüket betöltött középiskolások, egyetemisták, illetve pedagógiai szakszolgálati dolgozók voltak. A tipikus nyelvi fejlődést a gyermekcsoportban tesztekkel ellenőriztük (lásd alább), a felnőtt csoportban önbevallásra hagyatkoztunk; a személyeknek eszerint nem volt hallási, korábbi nyelvfejlődési problémájuk, illetve olyan beszédhibájuk sem, amit a logopédus képzettségű vizsgálatvezető észlelt volna. A kutatásban részt vevő gyermekek mindegyikének szüleitől írásos beleegyező nyilatkozatot kértünk. A kutatás tervét az ELTE Bárczi Gusztáv Gyógypedagógiai Karának Etikai Bizottsága előzetesen jóváhagyta.

Inger

Vizsgálatunkban kétféle módon szintetizált beszédingereket – mondatokat – használtunk: szinuszhullámú, illetve amplitúdómodulált beszédet.⁵ Ehhez egy előzetesen összeállított mondatlistáról hangfelvételt készítettünk; a mondatokat egy logopédus képzettségű férfi munkatárs mondta mikrofonba. A hangrögzítést a Bárczi Gusztáv Gyógypedagógiai Kar Ecséri úti hallásvizsgáló laborjában, tehát teljesen hangszige-

⁵ Egyik ingerként használt mondatunk természetes kiejtésben, valamint szinuszhullámú és amplitúdómodulált változatban is meghallgatható a következő internetes oldalon: <http://barczy.elte.hu/content/kutatas.t.1250>

telt körülmények között végeztük. Ehhez egy Audio Technica AT2010-es kondenzátor mikrofont használtuk (Focusrite scarlett 2i4 hangkártyával), amely elé pop filtert helyeztünk a torzítások kiküszöbölésére. A hangrögzítést az Audacity program vezérelte. A nyers felvételek hangerejét a Praat programmal egységesen 70 decibelre (az amplitúdó négyzetes középértéke: 0.0632 Pascal) állítottuk be, és ezen változathoz szintetizáltuk a szinuszhullámú és amplitúdómodulált mondatainkat.

A szinuszhullámú és amplitúdómodulált beszédet is e természetes kiejtés alapján készítettük. A szintézist a Praat szoftverre írt programokkal végeztük.⁶ E programok eredetijét Chris Darwin készítette;⁷ a jelen kutatásban csak kisebb módosításokat végeztünk rajtuk: az amplitúdómodulált beszéd frekvenciasávjait, illetve a szinuszbeszéd formánsainak keverési paramétereit módosítottuk. A szinuszhullámú szintézis az első három formánst (F1, F2, F3) helyettesítette egy-egy frekvencia- és amplitúdómodulált szinuszhullámmal, és ezeket adta össze.⁸ Az amplitúdómodulált beszédhangoknál hat sávot használtunk, miután a négy-sávú változat egy gyors előtesztben túl nehéznek tűnt a gyerekek számára. A hat frekvenciasáv (melyeken belül csak az amplitúdó változott, de az egyes sávoké egymástól függetlenül) a következő volt: 50–300 Hz; 300–600 Hz; 600–1200 Hz; 1200–2400 Hz; 2400–4800 Hz; 4800–8000 Hz. Minden, a vizsgálatban használt ingert .wav formátumban tároltunk.

A felhasznált mondatok listája a Függelékben található. Tíz alapmondatot használtunk, és mindegyiknek volt egy módosított szórendű (de azonos jelentésű), illetve egy jelentés szempontjából módosított változata is. A nyolc, éles próbában szereplő alapmondat két csoportra oszlott. Négy olyan alapmondat volt, melyben az ige argumentumainak felcserélésével (cselekvőváltással) kaptuk meg a szemantikai változatot (ezeket hívjuk „felcserélhető” vagy F típusú mondatoknak); a másik négy mondatban az igei csoport megváltoztatásával (tehát új kifejezés bevezetésével) módosítottuk a jelentést (NF típusú mondatok). A szórendi változatoknál a topikot változtattuk meg az alapmondathoz képest. A teljes ingerkészlet tehát 30 mondatból állt; ezek mindegyikéből szinuszhullámú, illetve amplitúdómodulált változatot is készítettünk. A mondatok hossza 9 és 18 szótag között változott; szerkesztésük fő szempontja az volt, hogy óvodás korú kísérleti személyeink számára jól érthetőek és szórakoztatóak legyenek, ezáltal segítve figyelmük fenntartását a kísérlet során.

⁶ A Praat hang-, illetve beszédelemző számítógépes rendszer készítői Paul Boersma és David Weenink (Phonetic Sciences, University of Amsterdam, Spuiistraat 210, 1012VT Amsterdam, Hollandia). A rendszer szabad felhasználású (<http://www.fon.hum.uva.nl/praat/>).

⁷ A szinuszhullámú beszédre használt eljárás forrása: http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praascripts/SWS; Az amplitúdómodulált beszéd szintézisé: http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praascripts/Shannon.

⁸ Egy lehetőség a szinuszhullámú beszédhangok nehezítésére, ha csak az alsó egy vagy két formáns szinuszváltozatát használjuk (F1, illetve F1+F2). A háromformánsos változat jelen vizsgálatban optimális választásnak tűnt a gyerekek és felnőttek teljesítményének összehasonlítására.

Kísérleti terv

A beszédészlelési vizsgálatban két gyakorló próba, majd ezt követően nyolc éles próba szerepelt, próbánként két szintetizált mondattal: egy alapmondattal, és annak egy módosított változatával. Minden éles próbában először egy alapmondat szintetizált változatát mutattuk be, és a személy elismételte azt, amit megértett belőle. Amennyiben nem volt tökéletes a mondat elismétlése, bemutattuk a mondatot természetes kiejtésben, és újra annak elismétlését kértük. Végül minden éles próbában egy módosított változatot – vagy a szemantikait, vagy a szórendit – mutattunk be szintetizált formában, és ennek elismétlését kértük. Amennyiben már az alapmondat első, szintetizált változatát tökéletesen elismételte a személy, a következő lépésben rögtön a módosított változatot adtuk, és a természetes kiejtést abban a próbában nem mutattuk be. A próbák ilyen felépítésével az volt a célunk, hogy a szinuszhullámú, illetve amplitúdómodulált beszéd észlelésénél tapasztalható „aha” élményt (tehát amikor a természetes kiejtés után újra halljuk a szintetizált változatot, és azt azonnal megértjük) egyfajta teljesítményfeladatba fordítsuk. A megváltozott jelentésű vagy szórendű mondat helyes elismétlése jelezheti azt, hogy a személy a természetes kiejtésből származó emlékein túl az alapmondat megváltozott jellegzetességeit is észleli, tehát egy kismértékű, de azonnali transzferhatás jelenik meg.⁹

Minden személynek a nyolc éles próbából négyben szemantikai (F vagy NF), a másik négyben pedig szórendi változatot adtunk az alapmondat után. Ezen kívül a kapott négy szemantikai változathoz mindenképpen kettő F típusú, kettő pedig NF típusú volt. Minden egyes csoportban (gyerek szinusz; gyerek VOC; felnőtt szinusz; felnőtt VOC) minden személy más és más mondat esetében kapott F, NF, illetve szórendi változatot, de e három típusba eső változatok száma mindenképpen ugyanannyi volt (2; 2; 4). A nyolc éles próba sorrendjét randomizáltuk; egy csoporton belül mindenki más sorrendben kapta az alapmondatokat. Ugyanakkor mind a négy csoport tagjai az ingerelrendezéseknek ugyanazt a halmazát kapták. Egészében tehát a módosított változatokkal együtt tizenhat mondat jelent meg a nyolc éles próbában, és hat a két gyakorló próbában (mivel itt a szórendi és az NF típusú szemantikai változatot is bemutattuk, mindkét esetben).

Eljárás és eszközök

Mivel célunk a tipikus nyelvfejlődésű populáció beszédészlelésének vizsgálata volt, ellenőriztük a tipikus, életkornak megfelelő nyelvi fejlettséget. A gyermekcsoportban ezt két módszeregyüttes használatával végeztük. Minden gyermek résztvevővel elvégeztük a SZÓL-E logopédiai szűrővizsgálatot (Kas, Lőrík, Bogáth és Mályi, 2012), valamint a

⁹ Sőt, a szemantikai, illetve szórendi változatok esetében e transzferhatás mértéke különbözhet is. Bár erre vonatkozóan részletes következtetést nem lehet levonni ilyen kis ingerkészlet (négy szemantikai és négy szórendi változat) alapján, amennyiben erre utaló jelet látunk, ez a későbbiekben alaposabban is megvizsgálható (Pléh és Sinkovics, 2011). Ugyanakkor a szinuszhullámú, illetve amplitúdómodulált mondatok hallgatása során fellépő gyors perceptuális tanulás hatását több más módon is lehet vizsgálni; az eredmények bemutatásánál ezt egy másik elemzési módszerrel is megtehetjük.

GOH-GMP teszt öt altesztjét: GMP1 – hallás; GMP2 – mondatazonosítás zajban; GMP3 – szóazonosítás zajban; GMP4 – szűk frekvenciás mondatok azonosítása; GMP5 – gyorsított mondatok azonosítása (Gósy, 1995). A felnőtt csoportban nem végeztünk nyelvi képességfelmérést; önbevallás alapján válogattuk a résztvevőket, akiknek eszerint nem volt hallási, korábbi nyelvfejlődési problémájuk és a logopédus képzettségű vizsgálatvezető által észlelhető beszédhibájuk sem. A gyerekeknél így két ülésben végeztük a vizsgálatot: az elsőben a nyelvi fejlettséget vizsgáltuk, a másodikban pedig a szintetizált beszédingereket mutattuk be. A felnőtt csoportokat egy ülésben vizsgáltuk.

A szinuszhullámú, illetve amplitúdómodulált beszédingereket egyforma eljárással mutattuk be. Az instrukcióban azt mondtuk a gyermek és felnőtt kísérleti személyeknek, hogy szokatlanul hangzó beszédet fognak hallani, mely első hallásra nem is hangzik beszédszerűen, de kis gyakorlás után megérthető. A szinuszhullámú mondatok esetében ezt madárcsicsergéshez, fűtyszóhoz, illetve a Csillagok háborúja c. filmből R2D2 beszédéhez hasonlítottuk; az amplitúdómodulált beszédet pedig a szél susogásához. Arra kértük a résztvevőket, hogy bármennyit is értenek meg a hallott beszédből, azt ismételjék meg. A hallott mondat elismétlését a természetes kiejtésű mondatok esetében is kértük. A válaszokat iPaden, Quick Voice programmal rögzítettük. Az ingereket hordozható számítógépről, egy Digital HD-838 típusú félig zárt fejhallgatón át mutattuk be, tehát a személyek fejhallgatóval a fejükön a vizsgálatvezető hangját is értették. Az ingerbemutató mindig csendes helyiségben történt. A hangerőt egy Voltcraft SL100 típusú hangnyomásmérővel 65 dB csúcsintenzitásra állítottuk be.

Az adatok értékelése

Minden mondat esetében a helyesen felismert szavak, szótagok, illetve magán- és más-salhangzók számát is rögzítettük. Elemzésünk alapja az, hogy a kísérleti személyek az egyes mondatoknak hány százalékát ismerik föl. Az angol nyelvi anyagon, szinuszhullámú, illetve amplitúdómodulált mondatokkal végzett vizsgálatokban általában a sikeresen felismert szavak és a mondat összes szavának arányát használják függő változóként (Rosner és mtsai, 2003; Nittrouer és mtsai, 2009; Nittrouer és Lowenstein, 2010; Lowenstein és mtsai, 2012; Hervais-Adelman és mtsai, 2011). Angol nyelvű ingeranyaggal végzett vizsgálatokban csak olyan esetekben alapozzák az elemzést szótagegységekre, amikor maguk az ingerek is szótagok, és nem szavak vagy mondatok (Nittrouer és Lowenstein, 2009; Hillenbrand és mtsai, 2011).¹⁰ Mi a felismert szótagok százalékos arányát használtuk; ennek indoklását lásd alább.

¹⁰ Egy további módszerben, melyet két év körüli gyerekekkel alkalmaztak, a kutatók egyszerű felszólító mondatokat játszottak le szinusz és VOC változatban (pl. „Nézd a kockát!”), és az ezek hatására előálló nézési időket mérték (minden felszólítás idején két kép volt jelen, melyek közül az egyiket szerepelt a megnevezett tárgy; Newman és mtsai, 2013).

EREDMÉNYEK

A felismerési mutatók összefüggése

Felmerült bennünk, hogy az angol és a magyar nyelv közti fonológiai különbségek miatt a felismert *szótagok* százalékos aránya a magyarban jobban értelmezhető mutató, mint az angol esetében, és valamivel finomabb felbontású is, mint a *szavak* felismerési aránya. Nem is szükségszerű – noha mondatok esetében igen valószínű –, hogy e két mutató szorosan korreláljon egymással. E kérdés megválaszolására egy előzetes vizsgálatot végeztünk: a nyolc, éles próbákban használt alapmondatra kiszámoltuk a két mutató (felismert szavak, illetve felismert szótagok százaléka) közötti korrelációt a teljes mintán. A korrelációs együtthatók értéke a nyolc alapmondatra 0.980 és 0.999 között változott. Mivel adataink jórészt nem normális eloszlásúak (lásd alább), a Spearman-féle rangkorrelációkat is kiszámoltuk; ezek értéke 0.976 és 0.999 között volt. Ez tehát arra utal, hogy ingereink esetében teljesen mindegy, hogy a szótagok vagy a szavak felismerési arányát tekintjük alapadatnak. A további elemzésekben a felismert *szótagok* százalékos arányait használjuk függő változóként.¹¹

A személyek nyolc alapmondataból számolt átlagos felismerési értéke a négy csoportból háromban normális volt, a felnőtt VOC csoportban a plafonhatásnak köszönhetően negatív ferdeség volt megfigyelhető (Shapiro-Wilk próba, $p < .05$; a ferdeségre: $z = -2.2275$, $p < .05$). Az egyes mondatok esetében a felismerési arányok eloszlása sehol nem volt normális a szélsőségek (0, illetve 1 közeli értékek) gyakori előfordulása miatt.

Személyek szerinti elemzések

Adataink elemzésénél először a kísérleti személyeket tekintettük individuumnak, és a teljesítményeiket elemeztük.

Átlagos felismerési teljesítmény a nyolc alapmondatra

Kísérleti személyeink, gyerekek és felnőttek egyaránt, jól teljesítettek mindkét inger-típus felismerésében. Ebben az első elemzésben a függő változó a személyek nyolc mondat alapján számolt átlagos teljesítménye volt (1. táblázat).

A felnőttek mindkét inger-típus mellett jobban teljesítettek, mint a gyerekek. Ezt két szempontos független mintás VA segítségével mutattuk meg (2 korcsoport x 2 inger-típus). Mindkét főhatás szignifikáns volt (korcsoport: $F(1,100) = 138.715$, $p < .0001$, $\eta^2 = 0.581$; inger-típus: $F(1,100) = 11.381$, $p < .001$, $\eta^2 = 0.102$). Utólag t-próbákkal is megvizsgáltuk, hogy az egyes inger-típusokra külön-külön is eltér-e a korcsoportok teljesítménye. Mind az SWS, mind az VOC inger-típusnál a felnőttek szignifikánsan jobban teljesítettek mint a gyerekek (SWS: $df = 47.03$, $t = 8.47$, $p < .001$, $r = 0.777$; VOC: $df =$

¹¹ A százalékos értékeket mindenhol 0 és 1 közé skáláztuk, tehát pl. a 0.54-es érték felel meg 54%-os felismerésnek.

1. táblázat. A szótagok átlagos felismerési aránya szórásokkal a nyolc mondatra, két csoportban és két ingerfeltétel (SWS, VOC) mellett

		Gyerek	Felnőtt
SWS	Átlag	0.504	0.788
	Szórás	0.135	0.104
VOC	Átlag	0.558	0.919
	Szórás	0.213	0.059

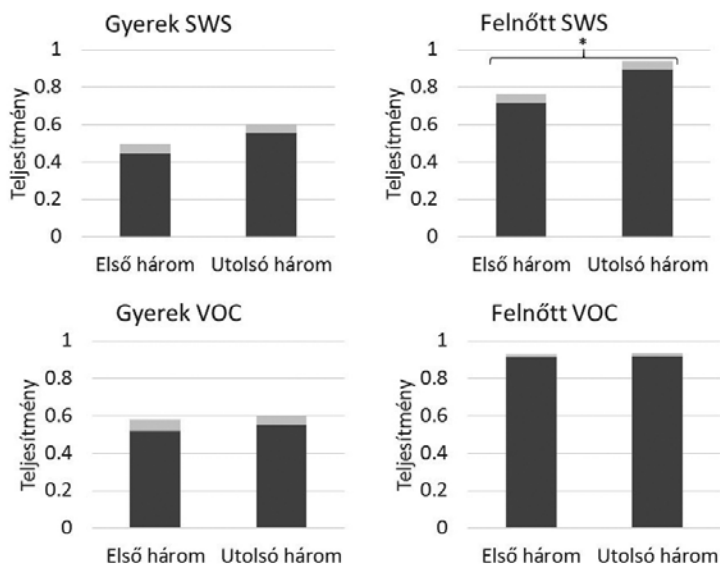
28.86, $t = 8.34$, $p < .001$, $r = 0.841$; Welch-próba). Ezenkívül a gyerekek egyformán teljesítettek a kétféle ingertípusra, a felnőttek viszont a VOC ingertípusra szignifikánsan jobban teljesítettek, mint az SWS-re ($df = 39.622$, $t = 5.58$, $p < .001$, $r = 0.663$; Welch-próba). Ennek ellenére a varianciaanalízisben az interakció nem volt szignifikáns, vagyis a korral való fejlődés mértéke adataink szerint nem különbözik a két ingertípus között. Ez az elemzés kiterjeszhető egy harmadik szempontra is, arra, hogy az egyes mondatok mennyire bizonyultak nehéznek a négy csoportban. Erre az elemzés második részében, a mondatok mint individuuumok vizsgálatakor térünk vissza.

A nemek közti különbséget ugyancsak a nyolc mondatból számolt átlagos felismerési arány alapján számoltuk. Háromszempontos, független mintás VA (2 nem X 2 életkor X 2 ingertípus) szignifikáns főhatást mutatott az életkorra ($p < .001$; $\eta^2 = 0.586$) és az ingertípusra ($p < .01$, $\eta^2 = 0.103$) is, nemekre azonban nem. Egyetlen interakció sem volt szignifikáns ebben a vizsgálatban. Ezután a nemek különbségét mind a négy csoportban külön-külön is megvizsgáltuk kétmintás t-próbával, és egyikben sem kaptunk szignifikáns különbséget. Így a további vizsgálatokban a ksz-ek nemét mint független változót nem használtuk.

Perceptuális tanulási hatások a próbák során

A próbák során megjelenő perceptuális tanulást először az alapmondatokon nyújtott teljesítmény segítségével vizsgáltuk. Ehhez minden személy esetében kiszámoltuk az általa kapott utolsó három (6., 7. és 8.), illetve első három mondaton mutatott felismerési teljesítmény különbségét (ezt nevezzük a továbbiakban U3-E3 változónak).¹² Ez a képzett különbségváltozó mind a négy csoportban normális eloszlású volt. Egymintás t-próbát végezve ezen a változón, a javulást (perceptuális tanulást) a 0-nál szignifikánsan magasabb érték jelenti: ebben az esetben az adott csoport személyei szignifikánsan jobban teljesítettek az utolsó három próbában kapott mondatokra, mint az első három próbában kapottakra. Ezt az összefüggést mind a négy csoportban megvizsgáltuk, és azt kaptuk, hogy egyedül a felnőtt SWS csoportban volt szignifikáns a javulás (egymintás t-próba, $t = 3.84$, $p < .001$, $r = 0.609$). A p érték a Bonferroni-korrektúra után

¹² Emlékezzünk rá, hogy a mondatok véletlen sorrendje miatt az egyes személyek más-más mondatot kaptak ugyanazon sorszámú próbában.



2. ábra. Teljesítményváltozás az első három, illetve utolsó három próba átlaga alapján a négy csoportban. A grafikonokon a fekete oszlopok az átlagokat, a világosszürke sávok a standard hibákat mutatják

($p_{\text{krit}} = 0.01/4 = 0.0025$) is szignifikáns maradt. Ezt a teljesítményváltozást mutatja a négy csoportban a 2. ábra.

A 2. ábrán az is látható, hogy a szinuszhullámú alapmondataink a gyerekek számára optimális nehézségűek voltak, és ugyanezek az ingerek még a felnőttek között is megfelelően diszkrimináltak. A hatsávós amplitúdómodulált ingerek a gyerekek számára szintén megfelelő nehézségűek voltak, a felnőtteknél azonban plafonhatást eredményeztek.

Érdeemes egy kicsit tovább vizsgálni a kérdést, hogy vajon a felnőtt SWS csoport, egyedül a négyből, tényleg mutatta-e jelét a perceptuális tanulásnak a nyolc próba során. Ehhez megvizsgáltunk még további három képzett változót: a nyolcadik, utolsó próbában mutatott teljesítmény és a legelső próbában mutatott teljesítmény különbségét (U1-E1), az utolsó kettő és első kettő különbségét (U2-E2), valamint az utolsó négy és az első négy eltérését (U4-E4). E próbák összegzett eredményeit mutatja a 2. táblázat.

A 2. táblázatból látható, hogy a felnőtt SWS csoportban a négy képzett változóból hárommal vizsgálva szignifikáns teljesítményjavulást kaptunk az első próbákról az utolsókra, és a negyedik képzett változó sem sokkal maradt el a szignifikáns szinttől. A másik három csoportban ellenben az elsőfajú hibakorrekció után egyetlen változó sem jelzett szignifikáns teljesítménynövekedést az ülés elejéről a végére. Ez az eredmény azt jelzi, hogy a jelen módszerrel vizsgálva az fSWS csoportban volt valamiféle teljesítménynövekedés, míg a másik három csoportban nem.

2. táblázat. A perceptuális tanulásra vonatkozó próbák összegző táblázata

	U1-E1	U2-E2	U3-E3	U4-E4
gySWS	NS	p<.05	NS	NS
fSWS	p < .01*	p < .01*	p < .001*	p < .0147
gyVOC	NS	NS	NS	NS
fVOC	NS	NS	NS	NS
$p_{krit} = 0.0125$	Wilcoxon		t-próba	

Megjegyzés: Csak a szignifikáns vagy ahhoz közeli értékeket írtuk ki. A négy oszlop a négy képzett változónak felel meg (U1-E1: utolsó próba mínusz első próba; U2-E2: utolsó két próba mínusz első két próba stb.); a sorok a négy vizsgálati csoportnak. A felnőtt SWS csoportban a négy változóból három szignifikáns teljesítményjavulást jelzett, egy pedig nem szignifikáns, de ahhoz közeli értéket adott. A másik három csoportban Bonferroni-korrekció után egyik változó sem jelzett szignifikáns javulást. A négy változó közül a normális eloszlásúakat t-próbával, az attól eltérőket Wilcoxon-próbával hasonlítottuk össze

Felmerül a kérdés, hogy mi okozhatta ezt a mintázatot. A 2. ábra jobb alsó grafikonján jól látható a felnőtt VOC csoportban megjelenő plafonhatás; e csoport tagjai már az első próbákban is majdnem tökéletesen teljesítettek, így az adott feltételek között nem tudtak további javulást mutatni az ülés végére. A bal oldali két grafikonon ellenben jól látszik, hogy a gyerekeknél nem volt plafonhatás, ennek ellenére mégsem javultak a viszonylag rövid kísérleti ülés alatt. További vizsgálatra érdemes kérdés, hogy az óvodás korú gyerekek kevésbé javulnak-e egy egyszeri kísérlet során a szintetizált mondatok felismerésében, mint a felnőttek; vagyis a gyors perceptuális tanulásban kevésbé hatékonyak-e, mint a felnőttek, hasonlóan a már említett Mooney-képekkel, tehát vizuális modalitásban végzett kevés létező vizsgálatok eredményéhez (lásd fent). Valamivel több próbából álló ülésekkel, és kicsit nehezebb (pl. hatsávós helyett négyávós) amplitúdómodulált ingerekkel ezt a hipotézist tovább lehet vizsgálni. A nehezebb VOC típusú ingerek várhatóan elmosás a felnőtteknél meglévő plafonhatást, míg a több próbát tartalmazó kísérletek kifejezettebbé tehetik a perceptuális tanulási hatást.

Mivel kiderült, hogy mondataink között viszonylag jelentős nehézségi különbségek vannak, és e különbségek összefüggenek a szintézis típusával is (lásd alább), felmerült a kérdés, hogy a perceptuális tanulással kapcsolatos eredményünket mennyire befolyásolhatták ezek a nehézségi különbségek. Lehet-e a perceptuális tanulásnak tűnő hatás valójában műtermék? Lehetséges-e például, hogy azok a személyek mutattak „teljesítménynövekedést”, akik az első próbákban nehezebb, míg az utolsókban könnyebb mondatokat kaptak?¹³ Ennek tisztázására a következő elemzést végeztük el. Először kiszámoltuk mind a nyolc alapmondatra és mind a négy csoportban a csoportra jellemző átlagos teljesítményt. Ezt egyfajta nehézségi mutatónak tekintettük: minél alacsonyabb volt az átlagos teljesítmény, annál nehezebb volt az adott mondat az adott csoport tagjainak. Ezután minden személyre az U3-E3 változó mellé (amely tehát a

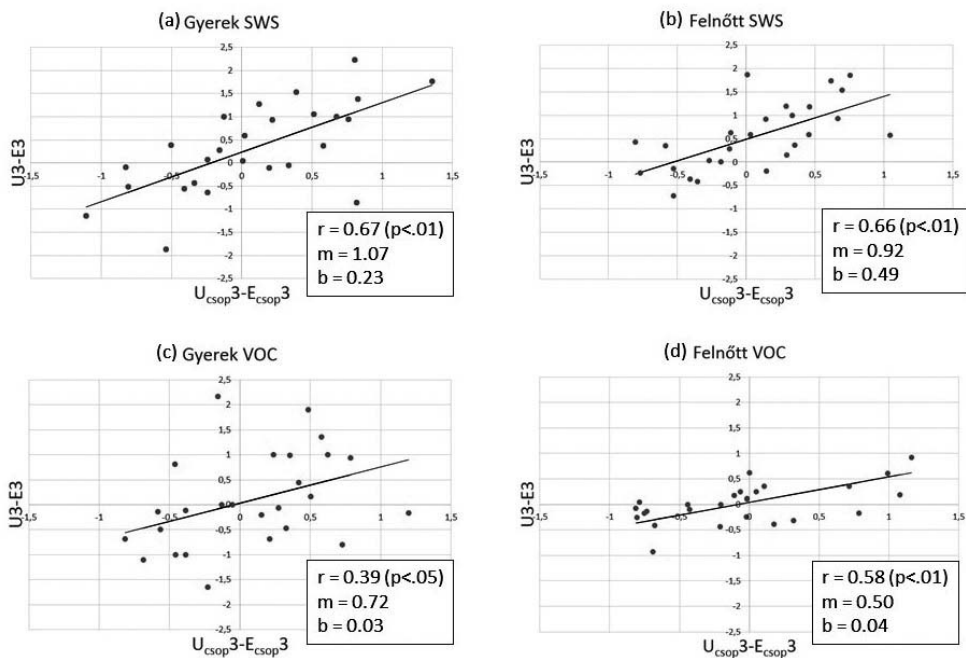
¹³ Megjegyezzük, hogy ez a jelenség nem okozhatott különbségeket a csoportok között a (látszólagos) perceptuális tanulásban, mert, mint fentebb leírtuk, mind a négy csoport tagjai ugyanabban a 26 pszeudo-random sorrendben kapták az ingereket.

személy saját teljesítményváltozását jellemzi) kiszámoltuk a megfelelő nehézségi indexet is: adott személy által az utolsó három próbában kapott mondatok nehézségeinek összege, mínusz az első három mondat nehézségösszege ($U_{csop} - 3 \cdot E_{csop}$). Ezután mind a négy csoportban kiszámoltuk a korrelációt és a regressziós egyenest is az U3-E3, illetve az $U_{csop} - 3 \cdot E_{csop}$ változók között (3. ábra). Az ábrán látható, hogy mind a négy csoportban azok teljesítménye javult az ülés elejéről a végére, akik az elején nehezebb, a végén pedig könnyebb mondatokat kaptak; ezt mutatja a mindegyik csoportban szignifikáns korreláció. A felnőtt SWS csoportban azonban e tényezőtől függetlenül is javult a teljesítmény az ülés elejéről a végére; ezt mutatja, hogy ebben a csoportban a legmagasabb a **b** érték (a regressziós egyenes függőleges tengellyel való metszéspontja), és ennek megfelelően az U3-E3 változó átlaga egyedül ebben a csoportban szignifikánsan nagyobb nullánál (lásd 2. táblázat). Az $U_{csop} - 3 \cdot E_{csop}$ változó átlaga egyik csoportban (így a felnőtt SWS-ben) sem tért el szignifikánsan a nullától.¹⁴ Összefoglalva, tehát ez az elemzés azt mutatta meg, hogy a kísérleti ülés során megfigyelhető teljesítményváltozásért egyrészt a személyek által kapott mondatok nehézsége (ennek a próbák során megjelenő változása) a felelős, másrészt ettől függetlenül kimutatható egy viszonylag gyors perceptuális tanulásként értelmezhető folyamat is, de csak a felnőtt SWS csoportban. További kutatásra érdemes kérdés, hogy vajon a felnőtt VOC csoportban is megfigyelhető-e ugyanez a hatás, amennyiben az ingerek nem túl könnyűek a személyek számára (azaz van hová javulniuk a kísérleti ülés elejéről a végére).

Kiegészítő elemzésként elvégeztünk egy-egy varianciaanalízist (2 ingertípus \times 2 életkor) az U3-E3, illetve az $U_{csop} - 3 \cdot E_{csop}$ változókra is. Utóbbi esetben, tehát a nehézségek változásánál, semmilyen szignifikáns hatást nem találtunk. Ez várható is volt, hiszen mind a négy csoportban a mondatoknak ugyanazt a 26 permutációját adtuk.¹⁵ Az U3-E3 változó varianciaanalízisében inger szerinti főhatást kaptunk: az SWS ingerekre magasabb volt a teljesítményváltozás értéke, mint a VOC változatokra ($p < 0.02$; $\eta^2 = 0.058$). A kor szerinti főhatás nem szignifikáns (az ülés elejéről a végére történt teljesítményváltozásban nem volt megfigyelhető egységes kor szerinti tendencia). Az interakció szintén nem volt szignifikáns. Ezt tovább elemezve azonban csak a felnőtt SWS és felnőtt VOC csoportok értékei különböztek szignifikánsan ($df = 34.5$; $t = 3.43$; $p < .002$, $r = 0.50$, Welch-próba), a gyerek SWS, illetve gyerek VOC csoportok különbsége nem volt szignifikáns. Ez megfelel annak a már bemutatott eredményünknek, hogy egyedül a felnőtt SWS csoportban volt az U3-E3 érték szignifikánsan nagyobb nullánál.

¹⁴ Az ábrán a meredekségek különbségéért jórészt az U3-E3 változó varianciájának csoportközi különbségei felelősek. Míg az $U_{csop} - 3 \cdot E_{csop}$ változó varianciája a négy csoportban 0.25 és 0.36 között változott, az U3-E3 változó varianciája a gyerek SWS csoportban 0.94, míg a felnőtt VOC csoportban 0.15 volt (ez az alacsony érték pedig a plafonhatás eredménye lehet).

¹⁵ Bár az egyes személyek eltértek abban, hogy az első három, illetve utolsó három próbában mely mondatokat kapták, s ennek megfelelően az egyes személyekhez tartozó $U_{csop} - 3 \cdot E_{csop}$ értékek is különböztek, a csoportok között ebben az indexben lényeges eltérést nem vártunk és nem is kaptunk. (Valamennyi eltérés adódhatott volna, mert egy adott mondat nehézségi indexe – átlagos felismerése – eltérhetett a négy csoportban.)



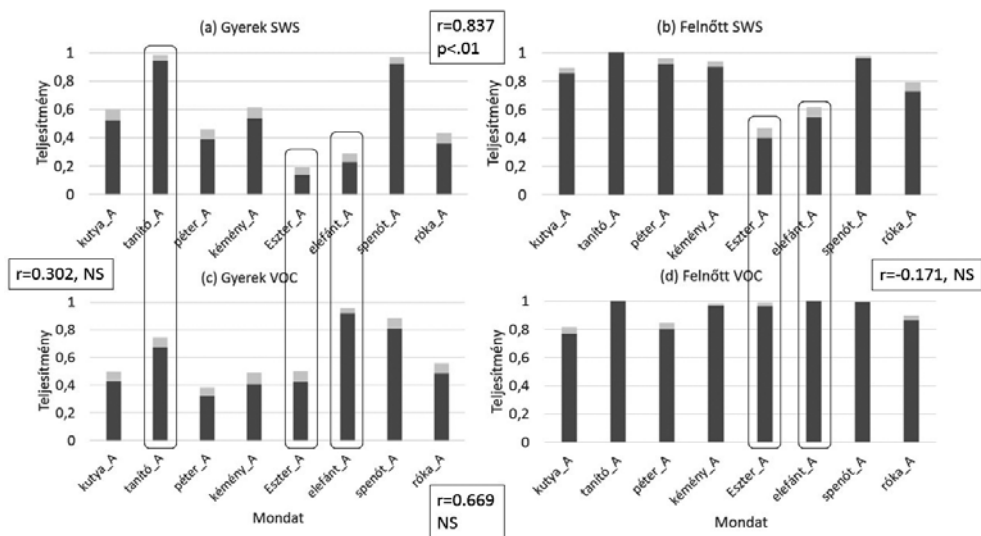
3. ábra. Az ingernehézség és a teljesítmény változása közötti összefüggés a négy csoportban A szövegdobozokban a korrelációs együttható (r), a regressziós egyenes meredeksége (m) és függőleges tengellyel való metszéspontja (b) látható. További magyarázat a szövegben

Mondatok szerinti elemzések

A négy csoport teljesítményei a nyolc alapmondaton

Mint korábban említettük (lásd A felismerési mutatók összefüggése c. részben), az egyes mondatok személyenkénti teljesítménymutatói nem normális eloszlásúak, de a nyolc mondaton nyújtott átlagos teljesítmény a négy csoportból háromban igen (a kivétel a felnőtt VOC csoport, és a normalitásvértés vélhető oka a plafonhatás). Ugyanakkor az egyes mondatok nem voltak egyformán nehezek a személyeink számára, sőt, ezek a nehézségi profilok különböztek a csoportok között (lásd 4. ábra). E mintázat elemzésére a legkézenfekvőbb egy háromszempontos VA (2 életkor X 2 ingertípus X 8 alapmondat; utóbbi ismétléses faktor) elvégzése. Ennek értelmezése azonban nem egyszerű, mert a normalitási feltétel mellett az ismétléses faktorra vonatkozó sphericitási és a csoportközi faktorokra vonatkozó szóráshomogenitási feltétel sem teljesült az adatainkra (a Maulchy-próba és a Levene-próba egyaránt szignifikáns volt). Ezt az elemzést mindazonáltal elvégeztük.¹⁶

¹⁶ Az eredményeket a következő módon látjuk értelmezhetőnek. A vegyes VA normalitás sértésére elfogadhatóan robusztus (Vargha, 2000, 421); a sphericitásvértéssel szemben a Huynh–Feldt, illetve Greenhouse–Geisser-eljárásokat használtuk. A csoportközi különbségekre vonatkozó robusztus eljárást ugyan nem számoltuk ki, de a csoportközi faktorokra lényegében ugyanazt kaptuk, mint az *Átlagos felismerési teljesítmény a nyolc alapmondatra* című részben leírt két szempontos (2 kor X 2 ingertípus) varianciaanalízisben. Az elsőfajú hibák valószínűségei pedig elég alacsonyak ahhoz, hogy viszonylag biztosak lehessünk a kapott hatások meglétében (lásd a főszöveget).



4. ábra. Az egyes alapmondatokra kapott nehézségek (átlagos felismerési teljesítmények) a négy csoportban. A kerettel jelölt mondatok azok, amelyek esetében a leglátványosabb volt a különbség az SWS és VOC változat felismerése között, adott korcsoporton belül. Ezt az öt esetet kétmintás t-próbával is megvizsgáltuk (a többit nem), és Bonferroni-korrekció után ($p_{krit} = 0.05/5 = 0.01$) valamennyi szignifikáns volt. Az ábrán látható korrelációs együtthatók a megfelelő két csoport teljesítményprofilja közötti összefüggést mutatják

Az ingertípus szerinti főhatás szignifikáns volt ($F(1,100) = 11.381$, $p < 0.01$, $\eta^2 = 0.102$), mert a VOC feltétel mellett jobban teljesítettek a személyek, mint SWS ingerekre (lásd 1. táblázat). A kor szerinti főhatás szintén szignifikáns volt ($F(1,100) = 138.715$, $p < 10^{-15}$, $\eta^2 = 0.58$), mert a felnőttek jobban teljesítettek a gyerekeknél. Ezek az eredmények egybevágnak az *Átlagos felismerési teljesítmény a nyolc alapmondatra* című részben bemutatott elemzéssel. A két csoportközi faktor közti interakció nem volt szignifikáns. Az ismétléses faktor szerinti főhatás is szignifikáns volt, azaz az egyes mondatok nem voltak egyforma nehezek ($F(7,700) = 34.040$, $p < 10^{-15}$, $\eta^2 = 0.254$). A mondatnehézségi profilok eltérnek az SWS, illetve VOC csoportban (gyerekek és felnőttek együtt; $F(7,700) = 26.57$, $p < 10^{-15}$, $\eta^2 = 0.21$), és a korcsoportok között is ($F(7,700) = 7.38$, $p < 10^{-7}$, $\eta^2 = 0.069$). A hármas interakció is szignifikáns volt, vagyis az életkor szerinti profilváltozás eltért a két ingertípus mellett ($F(7,700) = 3.02$, $p < .01$, $\eta^2 = 0.019$).

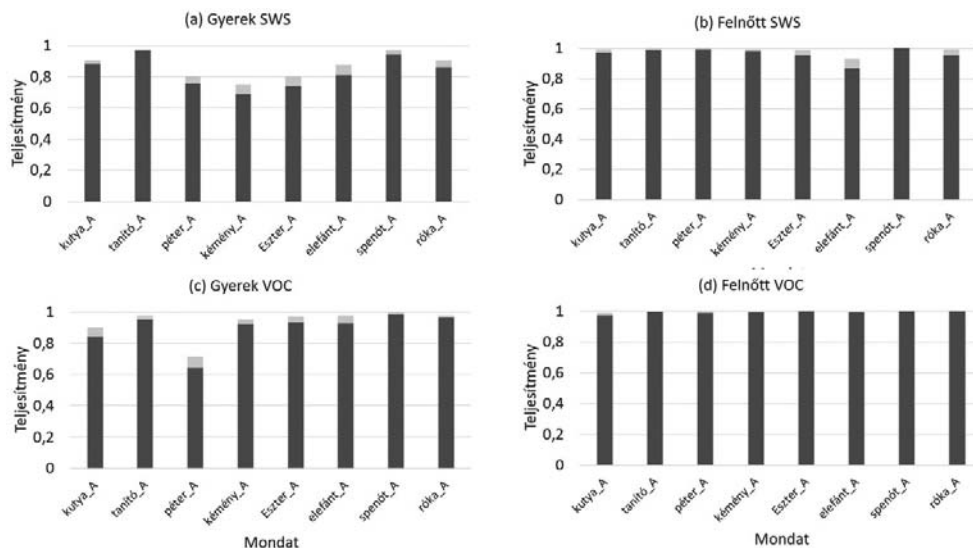
Kicsit másképpen is megvilágíthatjuk a számunkra releváns összefüggéseket. A gyerek SWS és a felnőtt SWS csoport tagjai nagyjából ugyanazokat a mondatokat ismerték fel könnyebben, illetve nehezebben ($r = 0.837$, $p < 0.01$; lásd 4. ábra). Hasonlót láthatunk a gyerek VOC és a felnőtt VOC csoportok viszonyában is, bár utóbbi esetben a korreláció csak közelítőleg volt szignifikáns ($p < 0.1$). Az alacsonyabb korrelációnak két oka lehet, az egyik a felnőtt VOC csoportban megfigyelhető plafonhatás, a másik pedig az alacsony szabadságfok ($df = 6$). Ugyanakkor sem a teljes gyerekcsoport két, ingertípus szerinti alcsoportjára, sem a két felnőtt alcsoportra nem volt szignifikáns a korreláció ($r = 0.302$, illetve $r = -0.171$). Úgy tűnik tehát, hogy a szintézis típusa

az elsődlegesen meghatározó tényező a mondatok egymáshoz viszonyított nehézsége szempontjából, és nem az életkor. Szorosabb ugyanis az összefüggés a szintézis típusa szerint, mint korcsoportok szerint. Ez a megfigyelés arra is utal, hogy mondataink nehézségére a jelentéstől független, akusztikai tulajdonságok, illetve ezek fonológiai megfeleltetése egyértelműen hatással volt. Más kérdés, hogy a gyermek–felnőtt teljesítménykülönbségre mely tényezők voltak elsősorban hatással: a jobb beszédészlelés vagy a nagyobb szókincs, esetleg mindkettő. E kérdés eldöntéséhez további vizsgálatokra van szükség.

Teljesítmény a módosított változatokon

Mondataink módosított változataiból az elvártnál sokkal kevesebb eredmény származott, és ennek oka az általános plafonhatás volt ezeken a mondatokon (5. ábra).

Minden csoport szignifikánsan jobban teljesített a módosított változatokon, mint az alapmondatokon (gyerek SWS: $t = 10.98$, $p < .001$, $r = 0.91$; felnőtt SWS: $t = 9.42$, $p < .001$, $r = 0.883$; gyerek VOC: $t = 9.53$, $p < .001$, $r = 0.886$; felnőtt VOC: $t = 6.85$, $p < .001$, $r = 0.808$). A további tendenciák megvilágításához kiszámoltuk minden személyre a módosított mondatokon produkált átlagos teljesítmény és az alapmondatokon adott átlagos teljesítmény különbségét. Mindenki jobban teljesített a módosított változatokon, mint az alapmondatokon, ezért ez az érték (nevezzük *javulási mutató*nak) mindenkinél pozitív volt. A javulási mutatón két szempontos (2 kor x 2 ingertípus) varianciaanalízist végeztünk. Itt az életkor szerinti főhatás szignifikáns volt ($F(1,100) = 66.545$, $p < 10^{-11}$, $\eta^2 = 0.4$). Ez az eredmény azonban a plafonhatásból eredő artefaktumnak tűnik. A már az alapmondatokon is jobban teljesítő felnőttek kisebbet tudtak javulni a



5. ábra. Teljesítmény a mondatok módosított változatán. Különösen a felnőtt csoportokban látható a plafonhatás, de a gyerekcsoportok is lényegesen jobban teljesítettek, mint az első bemutatásnál

módosított változatokon, mint az alapmondatokon gyengébb teljesítményt nyújtó gyerekek, ezért a gyerekek javulási mutatója volt szignifikánsan magasabb a felnőttekénél. Szintén a plafonhatásból eredhet a szignifikáns interakció is ($F(1,100) = 4.948, p < .03, \eta^2 = 0.047$): míg gyerekeknél SWS, illetve VOC ingerekre nem különbözött a javulási mutató értéke, addig felnőttek esetén SWS ingerekre nagyobb volt, mint VOC ingerekre. Itt is a 4. ábra két jobb oldali grafikonja segít az értelmezésben. A felnőtt VOC csoportban megfigyelhető a plafonhatás, a felnőtt SWS csoportban nem vagy sokkal kevésbé, ezért az SWS ingerek esetében biztosak lehetünk abban, hogy a korcsoportok közti teljesítménymelkedésnek volt elég tere.

Összefoglalva tehát, a javulási mutató e vizsgálatából egyetlen elméleti szempontból érdekes következtetést vonhatunk le. A személyek szubjektív, „aha” élménye azt jelzi, hogy a természetes kiejtés meghallgatása után ugyanazt a mondatot a megfelelően szintetizált változatban már tökéletesen megértik. Sőt, úgy tűnik, hogy az eredeti mondat módosított változatát is tökéletesen vagy közel tökéletesen értik szintetizált formában. A módosított változatokon tapasztalt igen magas teljesítmények miatt a szemantikai, illetve szórendi módosított változatok közti teljesítménykülönbségeket nem vizsgáltuk meg.

Az első főnévi csoport felismerésének hatása a mondat többi részének felismerésére

Ezt a vizsgálatot azért végeztük el, mert választ kerestünk arra, hogy mi okozhatja az egyes mondatok közötti felismerési, nehézségbeli különbségeket. Egyrészt azt láttuk, hogy az ingertípus igen erős hatást gyakorolt az egyes mondatok felismerésére, és ez a hatás jóval erősebb volt az életkori hatáznál (4. ábra). Kicsit pontosabban úgy fogalmazhatunk, hogy az életkor hatása a teljesítményre nem volt mondat-specifikus, ellenben a szintézis típusa mondat-specifikus hatást gyakorolt a teljesítményre. Ezzel párhuzamosan egy másik hatás is megfigyelhető a felismerésben, nevezetesen hogy a mondatok első főnévi csoportjának felismerése igen magasan korrelált az adott mondat maradék részének felismerésével. E hatás mögött vélhetőleg a szemantikai restauráció húzódnak meg. Egy konkrét példa erre a következő. Az „Eszter biciklit szeretne születésnapjára” mondat SWS változata a legnehezebb ingerünknek bizonyult. Itt a személyek (gyerekek és felnőttek is) gyakran „egyszer”-nek értették az első szót. Ez az értelmezés nyilván igencsak félreviszi a mondat többi részének dekódolását azáltal, hogy a mondat további tényleges szavainak feltételes valószínűségeit alaposan lecsökkenti. Bár előfordult olyan személy, aki az „Egyszer biciklit szeretne születésnapjára” mondattal válaszolt, a személyek többsége inkább értelmezési nehézséggel találta szemben magát, amit a töredékes válaszok nagyobb gyakorisága jelzett ennél a mondatnál. Egy ettől kissé különböző eset volt az „Az elefánt elbújt a fa mögé” mondat, itt ugyanis a személyek egy része láthatóan azon csodálkozott, hogy tud egy elefánt elbujni egy fa mögé, s ezért találhatták ezt a mondatot implauzibilis megoldásnak. Ennél a mondatnál tehát – legalábbis egyes személyeknél – szemantikai tényező is nehezíthette a felismerést. Megjegyzendő ugyanakkor, hogy csoportszinten a szintézis típusa is meghatározó tényezőnek tűnt az „elefántos” mondat felismerésében. E mondat ugyanis szinuszhullámú változatban nehezebb volt, mint amplitúdómoduláltban, és ezt a különbséget gyerekeknél, illetve felnőtteknél egyaránt megtaláltuk (4. ábra).

Ennél a mondatnál tehát egészében az lehetett a helyzet, hogy a hatsávós VOC változat könnyű érthetősége nagyjából minden szemantikai, értelmezési kétséget kizárt, így e feltétel mellett a gyerekek közel tökéletes, a felnőttek pedig kivétel nélkül tökéletes (100%-os) teljesítményt mutattak. A szinuszhullámú változat fonológiai dekódolhatósága viszont rosszabb volt, így itt tere nyílhatott a tűnődésnek, hogy mennyire lehet a hallott mondat értelmezése plauzibilis.

Az összes mondat alapján az is megfigyelhető volt, hogy a mondatok első főnévi csoportjának sikeres felismerése igen magasan korrelált az adott mondat többi részének felismerésével. A nyolc alapmondatra a négy csoportban a következő korrelációkat kaptuk (3. táblázat).

3. táblázat. Korrelációk a nyolc alapmondat első főnévi csoportjának, illetve az azt követő részének felismerése között

Csoport	r	p (df = 6)
gyerek SWS	0.959	p < .01
felnőtt SWS	0.836	p < .01
gyerek VOC	0.765	p < .05
felnőtt VOC	0.749	p < .05

További magyarázat a szövegben.

A 3. táblázat értelmezését segítő közöljük a gyerek SWS csoportban a nyolc alapmondatra vonatkozó részletes adatokat is (4. táblázat).

Ezen adatok alapján egy kétlépcsős modellt vázolhatunk fel a mondatok felismerésére vonatkozóan. Az első főnévi csoport felismerése, értelmezése során a legkisebb a lexikai restaurációs hatások tere, hiszen ekkor még a mondat elején járunk, a szövegkörnyezetből nem lehet következtetni az elhangzó jelek jelentésére. Ekkor a felismerésre az lehet döntő hatással, hogy a kétféle beszéd szintézis eredményeként előálló ingerek milyen mértékben segítik a megfelelő fonológiai kategóriák aktiválását. Feltehető például, hogy a szinuszhullámú beszéd a magánhangzók felismerését na-

4. táblázat. A gyerek SWS csoport adatai a nyolc alapmondat első főnévi csoportjának, illetve további részének felismeréséről

Eleje	Felismert%	Vége	Felismert%
A harapós kutya	0.545	elkergette a medvét a kertből	0.508
A tanítónéni	0.962	mindenkinek adott egy kis ajándékot	0.936
Péter	0.308	vendégül látta barátait a nagy sátorban	0.359
A kéményseprő	0.446	kölcsönadta a kabátját a kisfiúnak.	0.571
Eszter	0.077	biciklit szeretne születésnapjára	0.144
Az elefánt	0.058	elbújt a fa mögé	0.340
Minden gyerek	0.942	megette a spenótot	0.912
A ravasz róka	0.223	megszerezte a finom sajtot	0.436

A korrelációs együttható a két számszlop között ebben az esetben 0.959.

gyobb mértékben támogatja, mint a mássalhangzókét (Remez és Rubin, 1990; Remez és mtsai, 1994; Hillenbrand és mtsai, 2011). Ez azért lehet így, mert a frekvenciamodulált szinusz hullámok pontosan követik a magánhangzók formánsait, ellenben a mássalhangzók zörejkomponenseinek megfelelő frekvenciainformáció hiányzik belőlük. Az amplitúdómodulált beszéd ingerek viszont elsősorban a zöngétlen mássalhangzók zörejkomponenseit őrzik meg, a formánsokat nem. Egészében tehát a felismerés szintézisfüggőségét (az SWS és VOC közti különbséget) ez a tényező okozhatja, de, mivel mondatokról van szó, a következő szinten lexikai restaurációs hatások is jelentős szerepet játszanak, és ezek erősebben érvényesülnek mondataink predikátum tagjában, mint a szubjektumot leíró első főnévi csoportban. A sikerrel felismert első főnévi csoport pedig jó alapot ad a szemantikai restaurációra, miközben a személyek a mondat második részét hallgatják.

ÖSSZEFOGLALÁS ÉS MEGVITATÁS

Vizsgálatunkból a következő eredményeket érdemes kiemelni. Gyerekek és felnőttek egyaránt jól teljesítettek a szinusz hullámú, illetve amplitúdómodulált mondatok megértésében. Az amplitúdómodulált ingerek felnőttek számára könnyebbnek bizonyultak a szinusz hullámúaknál; gyerekeknél a kétféle ingertípus egyformán nehéz volt. A felnőttek mindkét ingertípusban a gyerekeknél jobban teljesítettek; a nemek között nem találtunk különbséget. Adatainkat összehasonlítva Nittrouer és munkatásai (Nittrouer és Lowenstein, 2010; Lowenstein és mtsai, 2012) eredményeivel, nem találtunk teljes összhangot. A két nyelv közötti összehasonlítást segíti az 5. táblázat.

A táblázat alapján a következő megfigyeléseket tehetjük.

(1) *A szinusz hullámú ingerek esetén az angol nyelvű vizsgálatban kapott teljesítmény magasabb az általunk megfigyeltnél, az amplitúdómoduláltak esetén ellenben alacsonyabb. Ez a kettős viszony attól függetlenül fennáll, hogy a saját adatainkból az átlagokat a 8 alapmondat avagy mind a 16 tesztmondat alapján számoljuk.* Szinusz hullámú ingerek esetén felnőtteknél és 7 éveseknél is olyan magas szinten van az angol személyek teljesítménye, amit mi a saját adatainkban plafonhatásnak minősítettünk (a felnőtt VOC csoportban). Az,

5. táblázat. Felismerési arányok Nittrouer és Lowenstein (2010) adatai, illetve a jelen vizsgálat alapján

	Nittrouer és Lowenstein, 2010			8 alapmondat		16 tesztmondat	
	Felnőttek	7 évesek	5 évesek	Felnőttek	5-6 évesek	Felnőttek	5-6 évesek
SWS	0.984 (0.014)	0.919 (0.03)	0.863 (0.061)	0.788 (0.104)	0.504 (0.135)	0.876 (0.120)	0.668 (0.203)
VOC	0.795 (0.083)	0.439 (0.166)	0.307 (0.179)	0.919 (0.059)	0.558 (0.213)	0.956 (0.057)	0.727 (0.238)

A cellákban a felső értékek átlagos felismerési arányt fejeznek ki, alattuk zárójelben a szórások. Nittrouer és Lowenstein az átlagosan felismert szavak számát közlik, ingertípusonként 6 gyakorló és 24 tesztmondat alapján. A saját adataink a felismert szótagok arányainak átlagai a 8 alapmondat, illetve mind a 16 tesztmondat alapján (tehát az alapmondatok módosított párját is beszámítva)

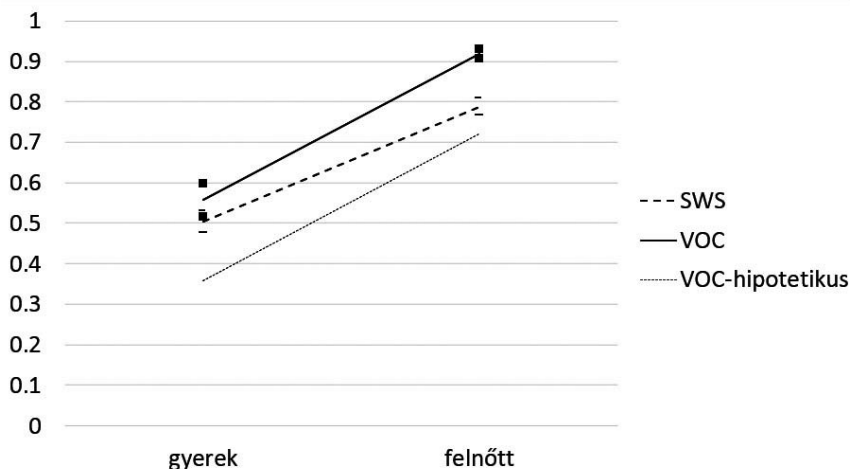
hogyan a VOC mondatokon a mi személyeink jobban teljesítettek az angolok alanyoknál, annak tulajdonítható, hogy míg Nittrouerek négysávós, mi hatsávós amplitúdómodulációt használtunk, és utóbbi megkönnyíti a beszédészlelést. A szinuszhullámú ingereken megfigyelhető ellentétes irányú különbség már nem értelmezhető ilyen könnyen, de elképzelhető, hogy az angol nyelvű vizsgálatban adott valamivel több gyakorlási lehetőség az egyik magyarázó tényező. Nittrouer és Lowenstein vizsgálatában a személyek ingertípusokként 6 gyakorló és 24 éles próbát kaptak. Emellett minden személy szinuszhullámú és amplitúdómodulált kísérleti blokkban is részt vett. A jelen vizsgálatban a szintézis típusa csoportközi változó volt, így minden személy egyetlen szintézistípusban összesen 20 szintetizált mondatot hallgatott meg. Nittrouerek személyeinek tehát egy ingertípuson belül is másfélszer annyi gyakorlásra volt lehetőségük, mint a mi vizsgálatunkban. A mi ingereink között az alapmondatok és módosított párjaik meglehetősen hasonlóak voltak egymáshoz. Ezért érdekesebb a 8 független alapmondaton alapuló átlagot használni az összehasonlításhoz. Nittrouerek vizsgálatában egy ingertípuson belül minden mondat független volt egymástól: nem voltak egymás transzformációi. Mindazonáltal a szinuszhullámú ingerekben megfigyelt nyelvközi teljesítménykülönbséget nem tekinthetjük tökéletesen megmagyarázottaknak, itt további vizsgálatokra van szükség.

(2) *Az angol nyelvű vizsgálatban minden korcsoportban jobban teljesítettek a személyek az SWS ingereken, mint a VOC ingereken. A mi adataink fordított viszonyt mutatnak.* E fordított viszony valószínű oka ismét az, hogy Nittrouerek négysávós, míg mi hatsávós amplitúdómodulációt használtunk. Mi a gyerekekkel végzett gyors elővizsgálatunk tapasztalatai alapján döntöttünk a hatsávós szintézis mellett: az első négy megvizsgált óvodás számára a négysávós ingerek frusztrálóan nehéznek tűntek, igen keveset értettek meg belőlük.

A Nittrouerek és saját adataink közti ellentmondás viszonylag egyszerűen feloldható, legalábbis egy kézenfekvő hipotézis segítségével. A 6. ábrát és az 5. táblázatot összevetve látható, hogy ha a VOC teljesítmény mindkét korcsoportunkban arányosan alacsonyabb lett volna, akkor Nittrouerek eredményeihez hasonló mintázat állna elő: mindkét korcsoport jobban teljesít az SWS mondatokra (az ábrán a hipotetikus VOC vonal az SWS alatt fut), de a gyerekeknél az SWS ingerek előnye nagyobb (hipotetikus) VOC ingerekhez képest, mint felnőtteknél. Lehetséges, hogy a négysávós VOC ingerekhez ragaszkodva ehhez hasonló eredményeket kaptunk volna; ezt további vizsgálatokkal kell tisztázni.

Egészében tehát nyilvánvaló, hogy az elszegényített beszédingerek észlelése jelentős mértékben fejlődik hatéves kortól felnőttkorig.

A rövid távú perceptuális tanulási jelenségekkel kapcsolatban a következőket figyelhetjük meg. A bevezetőben említett Mooney-képekkel összevetve az ezekkel kapcsolatban megfigyelhető azonnali perceptuális tanulás a szintetizált beszédingereknél is megjelent, ráadásul mindkét korcsoportban; ezt a módosított változatokon tapasztalt plafonhatás igazolja. Fontos kiemelni, hogy ez a hatás egyedi ingerekre vonatkozik, és nem tartalmaz utalást más, hasonló ingerekre való tanulási transzferre: egy inger elszegényített változatát először nem ismerjük fel, majd az eredeti, részletgazdag változat észlelése után azonnal felismerjük, és ezt egy jellegzetes „aha” élmény is kíséri. Itt tehát lényeges az ugrásszerű javulás. Ugyanakkor vannak jelei egy, a próbák során fokozato-



6. ábra. A szinuszhullámú, illetve hatsávós amplitúdómodulált ingerek teljesítményviszonya (felső két szakasz), illetve a négysávós amplitúdómodulált ingerekkel kapcsolatban feltételezett eredmények (legalsó szakasz). A valós adatoknál a standard hibákat is jelöltük. A hipotetikus vonal és a szaggatott vonal végpontjai közti különbségek hasonlítanak Nittrouer és munkatársai eredményeire

san kialakuló gyakorlási hatásnak is, tehát hogy új szintetizált mondatok észlelése az előzetes gyakorlás hatására fokozatosan javul. Ez a hatás már egyértelműen a tanulási transzferrel kapcsolatos, azonban csak felnőtteknél figyelhető meg, gyerekeknél nem. Sőt, a felnőttek is csak a szinuszhullámú ingerek esetén mutatták, a VOC ingereknél nem (valószínűleg a felnőtt VOC csoportban a plafonhatás miatt nem kaptuk meg a szignifikáns javulást). A szinuszhullámú ingereknél viszont érdekes, hogy a felnőtt csoportban már nyolc próba során is megfigyelhető a gyakorlási hatás; ennek kimutatása azonban összetett elemzést igényelt, melynek során a mondatok változó nehézségének hatását ki kellett szűrni. Ezt a rövid távú tanulási hatást Nittrouer és munkatársai nem vizsgálták, így ez – ismereteink szerint – új eredménynek tekinthető.

Mondataink nehézségével kapcsolatban megjegyzendő, hogy nem a szigorúan ki-egyenlített nehézségű ingerekre törekedtünk, hanem elsősorban arra, hogy a csak elnagyoltan hasonló szerkezetű mondatok az óvodás kísérleti személyeink figyelmét lekössék, bevonódásukat elősegítsék. A mondatnehézségek közti különbségeket és azok okait utólagos elemzéssel igyekeztünk föltárni. Ezzel kapcsolatban két megfigyelést fontos említeni. Egyrészt, mint a 4. ábrán látható, a mondatok nehézségi profilját alapvetően a szintézis típusa befolyásolta, és nem az életkor. Másrészt a mondatmegértés során fellépő szemantikai restaurációs hatások tetten érhetőek abban, hogy az első főnévi csoport értelmezésének sikere erősen összefügg a mondat többi részének megértésével.

További terveink között szerepel az ingerek alaposabb kalibrálása, a mondatok nyelvi nehézségének alaposabb kontrollja és valamivel több próba használata a kísér-

leti ülésekben. Ezekkel a módosításokkal a fent említett perceptuális tanulási hatást kívánjuk tisztább, egyértelműbb formában kimutatni. Ezen kívül a kisebb nyelvi egységek (szavak, álszavak, szótagok) vizsgálatát, illetve atipikus fejlődésű csoportok bevonását is tervezzük.

KÖSZÖNETNYILVÁNÍTÁS

A kézirat szerzői köszönetüket fejezik ki az ELTE Bárczi Gusztáv Gyógypedagógiai Kar munkatársainak, elsősorban Keresztessy Évának, a kar audioeszközeinek és hallásvizsgáló laborjának rendelkezésre bocsátásáért, valamint a Rákosmenti Mákvirág és a Rákosmenti Robogó óvoda vezetőinek, Horváth Máriának és Kiss Csabánének, a gyermek kísérleti személyek és a vizsgálatok helyszínének biztosításáért.

Függelék. A vizsgálatban használt ingerek

Tréning próbák alapmondatai	Szórendi változat	Szemantikai változat
A csibész kölyök betörte az ablakot a labdával. (NF)	Az ablakot a csibész kölyök betörte a labdával.	A jó gyerek betörte az ablakot a labdával.
A macska elbújt a nagymama ágyában. (NF)	A nagymama ágyában bújt el a macska.	A kutya bebújt a nagymama ágya alá.
Éles próbák alapmondatai	Szórendi változat	Szemantikai változat
A harapós kutya elkergette a medvét a kertből. (F)	A medvét a harapós kutya elkergette a kertből.	A harapós kutyát a medve elkergette a kertből.
A tanító néni mindenkinek adott egy kis ajándékot. (F)	Mindenkinek adott egy kis ajándékot a tanító néni.	A tanító néninek mindenki adott egy kis ajándékot.
Péter vendégül látta barátait a nagy sátorban. (F)	A nagy sátorban Péter vendégül látta a barátait.	A barátai vendégül látták Pétert a nagy sátorban.
A kéményseprő kölcsönadta a kabátját a kislának. (F)	A kabátját a kéményseprő kölcsönadta a kislának.	A kislány kölcsönadta a kabátját a kéményseprőnek.
Eszter biciklit szeretne születésnapjára. (NF)	Születésnapjára Eszter biciklit szeretne.	Eszter biciklit szeretne venni a testvérének.
Az elefánt elbújt a fa mögé. (NF)	A fa mögé bújt el az elefánt.	Az elefánt a fa mellett áll.
Minden gyerek megette a spenótot. (NF)	A spenótot minden gyerek megette.	Minden gyerek utálja a spenótot.
A ravasz róka megszerezte a finom sajtot. (NF)	A finom sajtot a ravasz róka szerezte meg.	A ravasz róka eldugta a finom sajtot.

FELHASZNÁLT IRODALOM

- Bent, T., Loebach, J., L., Phillips, L., & Pisoni, D., B. (2011). Perceptual adaptation to sinewave-vocoded speech across languages. *Journal of experimental psychology. Human perception and performance*, 37(5): 1607–1616.
- Bregman, A., S. (1990). *Auditory Scene Analysis*. Cambridge Mass: The MIT Press.
- Canepari, L. (2007). *Natural Phonetics and Tonetics*. LINCOM Publishers.
- Cherry, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975–979.
- Gósy M. (1995). GMP-Diagnosztika. A beszédészlelés és a beszédmegértés folyamatának vizsgálata. Budapest: NIKOL Kiadó.
- Gósy M. (2005). *Pszicholingvisztika*. Budapest: Osiris Kiadó.
- Hervais-Adelman, A., G., Davis, M., H., Johnsrude, I., S., Taylor, K., J., & Carlyon, R., P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of experimental psychology. Human perception and performance*, 37(1), 283–295.
- Hillenbrand, J., M., Clark, M., J., & Baer, C., A. (2011). Perception of sinewave vowels. *Journal of the Acoustical Society of America*, 129(6), 3991–4000.
- Honbolygó, F. (2007a). A hallási objektumok észlelése. In: Csépe V., Győri M. & Ragó A. (szerk.), *Általános pszichológia I. Észlelés és figyelem* (pp. 307–351). Budapest: Osiris Kiadó.
- Honbolygó, F. (2007b). A beszéd észlelése. In: Csépe V., Győri M. & Ragó A. (szerk.), *Általános pszichológia I. Észlelés és figyelem* (pp. 352–386). Budapest: Osiris Kiadó.
- Johnson, K. (2011). Speech perception. In: *Acoustic and Auditory Phonetics. 3rd edition* (pp. 118–150). Hoboken, NJ: Wiley-Blackwell.
- Kas B., Lőrík J., M. Bogáth R., Sz. Vékony A., & Mályi N. (2012). *SZÓL-E? Szűrőeljárás az óvodás-kori logopédiai ellátáshoz*. Budapest: Logotech Kiadó.
- Kovács, I., & Eisenberg, M. (2004) Human development of binocular rivalry. In: *Binocular Rivalry* (pp. 101–116). Cambridge Mass: The MIT Press.
- Ladefoged, P., & Broadbent, D., E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98–104.
- Lowenstein, J., H., Nittrouer, S., & Tarr, E. (2012). Children weight dynamic spectral structure more than adults: Evidence from equivalent signals. *Journal of the Acoustical Society of America*, 132(6), EL443–EL449.
- Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology*, 1, 219–228.
- Morrongiello, B., A., Robson, R., C., & Best, C., T. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, 37, 231–250.
- Newman, R., S., Chatterjee, M., Morini, G., & Nasuta, M. (2013). Toddlers' comprehension of noise-vocoded speech and sine-wave analogs to speech. *Proceedings of Meetings on Acoustics, Vol., 19* (ICA 2013, Montreal), <http://acousticalsociety.org/>
- Nilsson, M., Soli, S. D., & Gelnett, D. J. (1996). *Development and norming of a hearing in noise test for children*. Los Angeles, CA: House Ear Institute.
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085–1099.
- Nittrouer, S., & Lowenstein, J., H. (2009). Does harmonicity explain children's cue weighting of fricative-vowel syllables? *Journal of the Acoustical Society of America*, 125(3), 1679–1692.
- Nittrouer, S., & Lowenstein, J., H. (2010). Learning to perceptually organize speech signals in native fashion. *Journal of the Acoustical Society of America*, 127(3), 1624–1635.

- Nittrouer, S., Lowenstein, J., H., & Packer, R. (2009). Children discover the spectral skeletons in their native language before the amplitude envelopes. *Journal of Experimental Psychology Human Perception & Performance*, 35(4), 1245–1253.
- Pléh Cs., & Sinkovics B. (2011). A szórend és a fókuszbaemelés hatása a magyar mondatok emlékezeti leképezésében. *Magyar Pszichológiai Szemle*, 66(2), 321–333.
- Remez, R. (2005). Perceptual Organization of Speech. In: Pisoni, D. B., & Remez, R. E. (Eds), *Handbook of Speech Perception*. Oxford: Blackwell.
- Remez, R., E., & Rubin, P., E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics*, 48(4), 313–325.
- Remez, R., E., Rubin, P., E., Pisoni, D., B., & Carrell, T., D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947–950.
- Remez, R., E., Rubin, P., E., Nygaard, L., C., & Howell, W., A. (1987): Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology Human Perception & Performance*, 13(1) 40–61.
- Remez, R., E., Rubin, P., E., Berns, S., M., Pardo, J., S., & Lang, J., M. (1994): On the perceptual organization of speech. *Psychological Review*, 101(1), 129–156.
- Repp, B., H. (1982): Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81–110.
- Rosen, S., & Hui, S., N., C. (2015). Sine-wave and noise-vocoded sine-wave speech in a tone language: acoustic details matter. *Journal of the Acoustical Society of America*, 138, 3698–3702.
- Rosner, B., S., Talcott, J., B., Witton, C., Hogg, J., D., Richardson, A., J., Hanson, P., C., & Stein, J., F. (2003). The perception of "sine-wave speech" by adults with developmental dyslexia. *Journal of Speech, Language and Hearing Research*, 46, 68–79.
- Samuel, A., G., & Kat, D. (1996) Early levels of analysis of speech. *Journal of Experimental Psychology Human Perception & Performance*, 22(3), 676–694.
- Schwartz, J., Grimault N., Hupe, J., Moore, B., J., C., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B*, 367, 896–905.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Vargha A. (2000). *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Budapest: Pólya Kiadó.

PERCEPTION OF SINE-WAVE AND AMPLITUDE-MODULATED SENTENCES
IN HUNGARIANJAKAB, ZOLTÁN – NAGYNÉ RINGER, GABRIELLA – VÍG, JULIANNA –
SZABÓ, PÁL TAMÁS

Aims and scope: In the present study we used reduced speech stimuli to examine speech perception in 5-6-year-old children and adults.

Method: We used audiorecorded Hungarian sentences to synthesize sine-wave and vocoded versions of those sentences.

Results: Similarly to earlier developmental studies using English-language stimuli we found that both age groups performed well in understanding these forms of synthesized speech, and adults performed better than children. However, we found a pattern slightly different from the English data regarding the relation between SWS and VOC stimuli. While in studies using English stimuli children performed better at SWS than VOC stimuli, and the reverse relationship was observed in adults, we found equal performance for the two types of stimuli in children, and an advantage for VOC stimuli in adults. Lexical restoration was indicated by the finding that recognition of the first noun phrases in the sentences correlated strongly with recognition of the remaining part in both age groups, and both types of synthesis.

Conclusion: We reproduced some basic findings obtained in studies using English-language stimuli. As a novel finding we separated two types of perceptual learning that occurred in our sessions: the well-known type of instant perceptual learning, and a more gradual improvement over the course of the subsequent trials that reflected learning transfer between the stimuli. However, while instant perceptual learning was observable in both age groups, transfer-based improvement we only found in the adult group.

Keywords: Sine-wave speech, amplitude-modulated speech, speech perception, perceptual learning